

10/24/00
JC929 U.S. PTO

10-26-00

A

JC929 U.S. PTO
09/695812
10/24/00

1 IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

2 Inventorship..... Hunt et al.
3 Applicant..... Microsoft Corporation
4 Attorney's Docket No. MS1-547US
Title: System and Method for Distributed Management of Shared Computers

5 **TRANSMITTAL LETTER AND CERTIFICATE OF MAILING**

6 To: Commissioner of Patents and Trademarks,
Washington, D.C. 20231
7
8 From: Allan T. Sponseller (Tel. 509-324-9256; Fax 509-323-8979)
Lee & Hayes, PLLC
9 421 W. Riverside Avenue, Suite 500
Spokane, WA 99201

10 The following enumerated items accompany this transmittal letter and are being submitted for the
matter identified in the above caption.

- 11 1. Specification-title page, plus 52 pages, including 76 claims and Abstract
2. Transmittal letter including Certificate of Express Mailing
3. 7 Sheets Formal Drawings (Figs. 1-8)
4. Return Post Card

12 Large Entity Status [x] Small Entity Status []

13 Date: 10-24-2000

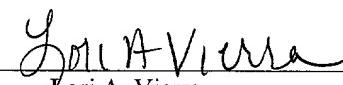
14 By:


15 Brian G. Hart
Reg. No. 44,421

16
17 **CERTIFICATE OF MAILING**
I hereby certify that the items listed above as enclosed are being deposited with the U.S. Postal
18 Service as either first class mail, or Express Mail if the blank for Express Mail No. is completed below, in
19 an envelope addressed to The Commissioner of Patents and Trademarks, Washington, D.C. 20231, on the
below-indicated date. Any Express Mail No. has also been marked on the listed items.

20 Express Mail No. (if applicable) E1685271215

21 Date: 10/24/00

22 By: 
Lori A. Vierra

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

**System and Method for Distributed Management of
Shared Computers**

Inventor(s):

Galen C. Hunt
Aamer Hydrie
Steven P. Levi
Jakob Rehof
David S. Stutz
Bassam Tabbara
Mark D. VanAntwerp
Robert V. Welland

ATTORNEY'S DOCKET NO. MS1-547US

1 **TECHNICAL FIELD**

2 This invention relates to computer system management. More particularly,
3 the invention relates to the distributed management of shared computers.

4

5 **BACKGROUND OF THE INVENTION**

6 The Internet and its use have expanded greatly in recent years, and this
7 expansion is expected to continue. One significant way in which the Internet is
8 used is the World Wide Web (also referred to as the "web"), which is a collection
9 of documents (referred to as "web pages") that users can view or otherwise render
10 and which typically include links to one or more other pages that the user can
11 access. Many businesses and individuals have created a presence on the web,
12 typically consisting of one or more web pages describing themselves, describing
13 their products or services, identifying other information of interest, allowing goods
14 or services to be purchased, etc.

15 Web pages are typically made available on the web via one or more web
16 servers, a process referred to as "hosting" the web pages. Sometimes these web
17 pages are freely available to anyone that requests to view them (e.g., a company's
18 advertisements) and other times access to the web pages is restricted (e.g., a
19 password may be necessary to access the web pages). Given the large number of
20 people that may be requesting to view the web pages (especially in light of the
21 global accessibility to the web), a large number of servers may be necessary to
22 adequately host the web pages (e.g., the same web page can be hosted on multiple
23 servers to increase the number of people that can access the web page
24 concurrently). Additionally, because the web is geographically distributed and has
25 non-uniformity of access, it is often desirable to distribute servers to diverse

1 remote locations in order to minimize access times for people in diverse locations
2 of the world. Furthermore, people tend to view web pages around the clock
3 (again, especially in light of the global accessibility to the web), so servers hosting
4 web pages should be kept functional 24 hours per day.

5 Managing a large number of servers, however, can be difficult. A reliable
6 power supply is necessary to ensure the servers can run. Physical security is
7 necessary to ensure that a thief or other mischievous person does not attempt to
8 damage or steal the servers. A reliable Internet connection is required to ensure
9 that the access requests will reach the servers. A proper operating environment
10 (e.g., temperature, humidity, etc.) is required to ensure that the servers operate
11 properly. Thus, "co-location facilities" have evolved which assist companies in
12 handling these difficulties.

13 A co-location facility refers to a complex that can house multiple servers.
14 The co-location facility typically provides a reliable Internet connection, a reliable
15 power supply, and proper operating environment. The co-location facility also
16 typically includes multiple secure areas (e.g., cages) into which different
17 companies can situate their servers. The collection of servers that a particular
18 company situates at the co-location facility is referred to as a "server cluster", even
19 though in fact there may only be a single server at any individual co-location
20 facility. The particular company is then responsible for managing the operation of
21 the servers in their server cluster.

22 Such co-location facilities, however, also present problems. One problem
23 is data security. Different companies (even competitors) can have server clusters
24 at the same co-location facility. Care is required, in such circumstances, to ensure
25 that data received from the Internet (or sent by a server in the server cluster) that is

1 intended for one company is not routed to a server of another company situated at
2 the co-location facility.

3 An additional problem is the management of the servers once they are
4 placed in the co-location facility. Currently, a system administrator from a
5 company is able to contact a co-location facility administrator (typically by
6 telephone) and ask him or her to reset a particular server (typically by pressing a
7 hardware reset button on the server, or powering off then powering on the server)
8 in the event of a failure of (or other problem with) the server. This limited reset-
9 only ability provides very little management functionality to the company.
10 Alternatively, the system administrator from the company can physically travel to
11 the co-location facility him/her-self and attend to the faulty server. Unfortunately,
12 a significant amount of time can be wasted by the system administrator in
13 traveling to the co-location facility to attend to a server. Thus, it would be
14 beneficial to have an improved way to manage remote server computers at a co-
15 location facility.

16 Another problem concerns the enforcement of the rights of both the
17 operators of the servers in the co-location facility and the operators of the web
18 service hosted on those servers. The operators of the servers need to be able to
19 maintain their rights (e.g., re-possessing areas of the facility where servers are
20 stored), even though the servers are owned by the operators of the web service.
21 Additionally, the operators of the web service need to be assured that their data
22 remains secure.

23 The invention described below addresses these disadvantages, improving
24 the distributed management of shared computers in co-location facilities.
25

1 **SUMMARY OF THE INVENTION**

2 Distributed management of shared computers is described herein.

3 According to one aspect, a multi-tiered management architecture is
4 employed including an application development tier, an application operations tier,
5 and a cluster operations tier. In the application development tier, applications are
6 developed for execution on one or more server computers. In the application
7 operations tier, execution of the applications is managed and sub-boundaries
8 within a cluster of servers at a co-location facility may be established. In the
9 cluster operations tier, operation of the server computers is managed without
10 concern for what applications are executing on the one or more server computers,
11 and server cluster boundaries at the co-location facility may be established.

12 According to another aspect, a co-location facility includes multiple server
13 clusters, each corresponding to a different customer. For each server cluster, a
14 cluster operations management console is implemented locally at the co-location
15 facility to manage hardware operations of the cluster, and an application
16 operations management console is implemented at a location remote from the co-
17 location facility to manage software operations of the cluster. In the event of a
18 hardware failure, the cluster operations management console takes corrective
19 action (e.g., notifying an administrator at the co-location facility or attempting to
20 correct the failure itself). In the event of a software failure, the application
21 operations management console takes corrective action (e.g., notifying one of the
22 customer's administrators or attempting to correct the failure itself).

23 According to another aspect, boundaries of a server cluster are established
24 by a cluster operations management console. Establishment of the boundaries
25 ensures that data is routed only to nodes within the server cluster, and not to other

1 nodes at the co-location facility that are not part of the server cluster. Further sub-
2 boundaries within a server cluster may be established by an application operations
3 management console to ensure data is routed only to particular nodes within the
4 server cluster.

5 According to another aspect, rights to multiple server computers to be
6 located at a co-location facility are sold to a customer and a multiple-tiered
7 management scheme is enforced on the server computers. According to the
8 multiple-tiered management scheme, hardware operation of the server computers
9 is managed locally at the co-location facility whereas software operation of the
10 server computers is managed from a location remote from the co-location facility.
11 The server computers can be either sold to the customer or leased to the customer.

12 According to another aspect, a landlord/tenant relationship is created using
13 one or more server computers at a co-location facility. The operator of the co-
14 location facility supplies the facility as well as the servers (and thus can be viewed
15 as a "landlord"), while customers of the facility lease the use of the facility as well
16 as servers at that facility (and thus can be viewed as "tenants"). This
17 landlord/tenant relationship allows the landlord to establish clusters of computers
18 for different tenants and establish boundaries between clusters so that a tenant's
19 data does not pass beyond its cluster (and to another tenant's cluster). Additionally,
20 encryption is employed in various manners to assure the tenant that
21 information stored at the servers it leases cannot be viewed by anyone else, even if
22 the tenant terminates its lease or returns to the landlord one of the servers it is
23 leasing.

24 According to another aspect, a multi-tiered management architecture is
25 employed in managing computers that are not part of a co-location facility. This

1 multi-tiered architecture is used for managing computers (whether server
2 computers or otherwise) in a variety of settings, such as businesses, homes, etc.

3

4 **BRIEF DESCRIPTION OF THE DRAWINGS**

5 The present invention is illustrated by way of example and not limitation in
6 the figures of the accompanying drawings. The same numbers are used
7 throughout the figures to reference like components and/or features.

8 Fig. 1 shows a client/server network system and environment such as may
9 be used with certain embodiments of the invention.

10 Fig. 2 shows a general example of a computer that can be used in
11 accordance with certain embodiments of the invention.

12 Fig. 3 is a block diagram illustrating an exemplary co-location facility in
13 more detail.

14 Fig. 4 is a block diagram illustrating an exemplary multi-tiered
15 management architecture.

16 Fig. 5 is a block diagram illustrating an exemplary node in more detail in
17 accordance with certain embodiments of the invention.

18 Fig. 6 is a flowchart illustrating an exemplary process for encryption key
19 generation and distribution in accordance with certain embodiments of the
20 invention.

21 Fig. 7 is a flowchart illustrating an exemplary process for the operation of a
22 cluster operations management console in accordance with certain embodiments
23 of the invention.

1 Fig. 8 is a flowchart illustrating an exemplary process for the operation of
2 an application operations management console in accordance with certain
3 embodiments of the invention.

4

5 **DETAILED DESCRIPTION**

6 Fig. 1 shows a client/server network system and environment such as may
7 be used with certain embodiments of the invention. Generally, the system
8 includes multiple (*n*) client computers 102 and multiple (*m*) co-location facilities
9 104 each including multiple clusters of server computers (server clusters) 106.
10 The servers and client computers communicate with each other over a data
11 communications network 108. The communications network in Fig. 1 comprises a
12 public network 108 such as the Internet. Other types of communications networks
13 might also be used, in addition to or in place of the Internet, including local area
14 networks (LANs), wide area networks (WANs), etc. Data communications
15 network 108 can be implemented in any of a variety of different manners,
16 including wired and/or wireless communications media.

17 Communication over network 108 can be carried out using any of a wide
18 variety of communications protocols. In one implementation, client computers
19 102 and server computers in clusters 106 can communicate with one another using
20 the Hypertext Transfer Protocol (HTTP), in which web pages are hosted by the
21 server computers and written in a markup language, such as the Hypertext Markup
22 Language (HTML) or the eXtensible Markup Language (XML).

23 In the discussions herein, embodiments of the invention are described
24 primarily with reference to implementation at a co-location facility (such as
25 facility 104). The invention, however, is not limited to such implementations and

can be used for distributed management in any of a wide variety of situations. For example, in situations where all of the servers at a facility are owned or leased to the same customer, in situations where a single computing device (e.g., a server or client) is being managed, in situations where computers (whether servers or otherwise) in a business or home environment are being managed, etc.

In the discussion herein, embodiments of the invention are described in the general context of computer-executable instructions, such as program modules, being executed by one or more conventional personal computers. Generally, program modules include routines, programs, objects, components, data structures, etc. that perform particular tasks or implement particular abstract data types. Moreover, those skilled in the art will appreciate that various embodiments of the invention may be practiced with other computer system configurations, including hand-held devices, gaming consoles, Internet appliances, multiprocessor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, and the like. In a distributed computer environment, program modules may be located in both local and remote memory storage devices.

Alternatively, embodiments of the invention can be implemented in hardware or a combination of hardware, software, and/or firmware. For example, all or part of the invention can be implemented in one or more application specific integrated circuits (ASICs) or programmable logic devices (PLDs).

Fig. 2 shows a general example of a computer 142 that can be used in accordance with certain embodiments of the invention. Computer 142 is shown as an example of a computer that can perform the functions of a client computer 102 of Fig. 1, a computer or node in a co-location facility 104 of Fig. 1 or other

1 location (e.g., node 248 of Fig. 5 below), or a local or remote management console
2 as discussed in more detail below.

3 Computer 142 includes one or more processors or processing units 144, a
4 system memory 146, and a bus 148 that couples various system components
5 including the system memory 146 to processors 144. The bus 148 represents one
6 or more of any of several types of bus structures, including a memory bus or
7 memory controller, a peripheral bus, an accelerated graphics port, and a processor
8 or local bus using any of a variety of bus architectures. The system memory
9 includes read only memory (ROM) 150 and random access memory (RAM) 152.
10 A basic input/output system (BIOS) 154, containing the basic routines that help to
11 transfer information between elements within computer 142, such as during start-
12 up, is stored in ROM 150.

13 Computer 142 further includes a hard disk drive 156 for reading from and
14 writing to a hard disk, not shown, connected to bus 148 via a hard disk driver
15 interface 157 (e.g., a SCSI, ATA, or other type of interface); a magnetic disk drive
16 158 for reading from and writing to a removable magnetic disk 160, connected to
17 bus 148 via a magnetic disk drive interface 161; and an optical disk drive 162 for
18 reading from or writing to a removable optical disk 164 such as a CD ROM, DVD,
19 or other optical media, connected to bus 148 via an optical drive interface 165.
20 The drives and their associated computer-readable media provide nonvolatile
21 storage of computer readable instructions, data structures, program modules and
22 other data for computer 142. Although the exemplary environment described
23 herein employs a hard disk, a removable magnetic disk 160 and a removable
24 optical disk 164, it should be appreciated by those skilled in the art that other types
25 of computer readable media which can store data that is accessible by a computer,

1 such as magnetic cassettes, flash memory cards, digital video disks, random access
2 memories (RAMs) read only memories (ROM), and the like, may also be used in
3 the exemplary operating environment.

4 A number of program modules may be stored on the hard disk, magnetic
5 disk 160, optical disk 164, ROM 150, or RAM 152, including an operating system
6 170, one or more application programs 172, other program modules 174, and
7 program data 176. A user may enter commands and information into computer
8 142 through input devices such as keyboard 178 and pointing device 180. Other
9 input devices (not shown) may include a microphone, joystick, game pad, satellite
10 dish, scanner, or the like. These and other input devices are connected to the
11 processing unit 144 through an interface 168 that is coupled to the system bus. A
12 monitor 184 or other type of display device is also connected to the system bus
13 148 via an interface, such as a video adapter 186. In addition to the monitor,
14 personal computers typically include other peripheral output devices (not shown)
15 such as speakers and printers.

16 Computer 142 optionally operates in a networked environment using
17 logical connections to one or more remote computers, such as a remote computer
18 188. The remote computer 188 may be another personal computer, a server, a
19 router, a network PC, a peer device or other common network node, and typically
20 includes many or all of the elements described above relative to computer 142,
21 although only a memory storage device 190 has been illustrated in Fig. 2. The
22 logical connections depicted in Fig. 2 include a local area network (LAN) 192 and
23 a wide area network (WAN) 194. Such networking environments are
24 commonplace in offices, enterprise-wide computer networks, intranets, and the
25 Internet. In the described embodiment of the invention, remote computer 188

1 executes an Internet Web browser program (which may optionally be integrated
2 into the operating system 170) such as the "Internet Explorer" Web browser
3 manufactured and distributed by Microsoft Corporation of Redmond, Washington.

4 When used in a LAN networking environment, computer 142 is connected
5 to the local network 192 through a network interface or adapter 196. When used
6 in a WAN networking environment, computer 142 typically includes a modem 198
7 or other component for establishing communications over the wide area network
8 194, such as the Internet. The modem 198, which may be internal or external, is
9 connected to the system bus 148 via an interface (e.g., a serial port interface 168).
10 In a networked environment, program modules depicted relative to the personal
11 computer 142, or portions thereof, may be stored in the remote memory storage
12 device. It is to be appreciated that the network connections shown are exemplary
13 and other means of establishing a communications link between the computers
14 may be used.

15 Generally, the data processors of computer 142 are programmed by means
16 of instructions stored at different times in the various computer-readable storage
17 media of the computer. Programs and operating systems are typically distributed,
18 for example, on floppy disks or CD-ROMs. From there, they are installed or
19 loaded into the secondary memory of a computer. At execution, they are loaded at
20 least partially into the computer's primary electronic memory. The invention
21 described herein includes these and other various types of computer-readable
22 storage media when such media contain instructions or programs for implementing
23 the steps described below in conjunction with a microprocessor or other data
24 processor. The invention also includes the computer itself when programmed
25 according to the methods and techniques described below. Furthermore, certain

1 sub-components of the computer may be programmed to perform the functions
2 and steps described below. The invention includes such sub-components when
3 they are programmed as described. In addition, the invention described herein
4 includes data structures, described below, as embodied on various types of
5 memory media.

6 For purposes of illustration, programs and other executable program
7 components such as the operating system are illustrated herein as discrete blocks,
8 although it is recognized that such programs and components reside at various
9 times in different storage components of the computer, and are executed by the
10 data processor(s) of the computer.

11 Fig. 3 is a block diagram illustrating an exemplary co-location facility in
12 more detail. Co-location facility 104 is illustrated including multiple nodes (also
13 referred to as server computers) 210. Co-location facility 104 can include any
14 number of nodes 210, and can easily include an amount of nodes numbering into
15 the thousands.

16 The nodes 210 are grouped together in clusters, referred to as server
17 clusters (or node clusters). For ease of explanation and to avoid cluttering the
18 drawings, only a single cluster 212 is illustrated in Fig. 3. Each server cluster
19 includes nodes 210 that correspond to a particular customer of co-location facility
20 104. The nodes 210 of a server cluster are physically isolated from the nodes 210
21 of other server clusters. This physical isolation can take different forms, such as
22 separate locked cages or separate rooms at co-location facility 104. Physically
23 isolating server clusters ensures customers of co-location facility 104 that only
24 they can physically access their nodes (other customers cannot). Alternatively,

1 server clusters may be logically, but not physically, isolated for each other (e.g.,
2 using cluster boundaries as discussed in more detail below).

3 A landlord/tenant relationship (also referred to as a lessor/lessee
4 relationship) can also be established based on the nodes 210. The owner (and/or
5 operator) of co-location facility 104 owns (or otherwise has rights to) the
6 individual nodes 210, and thus can be viewed as a "landlord". The customers of
7 co-location facility 104 lease the nodes 210 from the landlord, and thus can be
8 viewed as a "tenant". The landlord is typically not concerned with what types of
9 data or programs are being stored at the nodes 210 by the tenant, but does impose
10 boundaries on the clusters that prevent nodes 210 from different clusters from
11 communicating with one another, as discussed in more detail below.

12 The landlord/tenant relationship is discussed herein primarily with
13 reference to only two levels: the landlord and the tenant. However, in alternate
14 embodiments this relationship can be expanded to any number of levels. For
15 example, the landlord may share its management responsibilities with one or more
16 sub-landlords (each of which would have certain managerial control over one or
17 more nodes 210), and the tenant may similarly share its management
18 responsibilities with one or more sub-tenants (each of which would have certain
19 managerial control over one or more nodes 210).

20 Although physically isolated, nodes 210 of different clusters are often
21 physically coupled to the same transport medium (or media) 211 that enables
22 access to network connection(s) 216, and possibly application operations
23 management console 242, discussed in more detail below. This transport medium
24 can be wired or wireless.

As each node 210 can be coupled to a shared transport medium 211, each node 210 is configurable to restrict which other nodes 210 data can be sent to or received from. Given that a number of different nodes 210 may be included in a tenant's server cluster, the tenant may want to be able to pass data between different nodes 210 within the cluster for processing, storage, etc. However, the tenant will typically not want data to be passed to other nodes 210 that are not in the server cluster. Configuring each node 210 in the cluster to restrict which other nodes 210 data can be sent to or received from allows a boundary for the server cluster to be established and enforced. Establishment and enforcement of such server cluster boundaries prevents tenant data from being erroneously or improperly forwarded to a node that is not part of the cluster.

These initial boundaries established by the landlord prevent communication between nodes 210 of different tenants, thereby ensuring that each tenant's data can be passed to other nodes 210 of that tenant. The tenant itself may also further define sub-boundaries within its cluster, establishing sub-clusters of nodes 210 that data cannot be communicated out of (or in to) either to or from other nodes in the cluster. The tenant is able to add, modify, remove, etc. such sub-cluster boundaries at will, but only within the boundaries defined by the landlord (that is, the cluster boundaries). Thus, the tenant is not able to alter boundaries in a manner that would allow communication to or from a node 210 to extend to another node 210 that is not within the same cluster.

Co-location facility 104 supplies reliable power 214 and reliable network connection(s) 216 to each of the nodes 210. Power 214 and network connection(s) 216 are shared by all of the nodes 210, although alternatively separate power 214 and network connection(s) 216 may be supplied to nodes 210 or groupings (e.g.,

1 clusters) of nodes. Any of a wide variety of conventional mechanisms for
2 supplying reliable power can be used to supply reliable power 214, such as power
3 received from a public utility company along with backup generators in the event
4 of power failures, redundant generators, batteries, fuel cells, or other power
5 storage mechanisms, etc. Similarly, any of a wide variety of conventional
6 mechanisms for supplying a reliable network connection can be used to supply
7 network connection(s) 216, such as redundant connection transport media,
8 different types of connection media, different access points (e.g., different Internet
9 access points, different Internet service providers (ISPs), etc.).

10 In certain embodiments, nodes 210 are leased or sold to customers by the
11 operator or owner of co-location facility 104 along with the space (e.g., locked
12 cages) and service (e.g., access to reliable power 214 and network connection(s)
13 216) at facility 104. In other embodiments, space and service at facility 104 may
14 be leased to customers while one or more nodes are supplied by the customer.

15 Management of each node 210 is carried out in a multiple-tiered manner.
16 Fig. 4 is a block diagram illustrating an exemplary multi-tiered management
17 architecture. The multi-tiered architecture includes three tiers: a cluster
18 operations management tier 230, an application operations management tier 232,
19 and an application development tier 234. Cluster operations management tier 230
20 is implemented locally at the same location as the server(s) being managed (e.g., at
21 a co-location facility) and involves managing the hardware operations of the
22 server(s). In the illustrated example, cluster operations management tier 230 is not
23 concerned with what software components are executing on the nodes 210, but
24 only with the continuing operation of the hardware of nodes 210 and establishing
25 any boundaries between clusters of nodes.

1 The application operations management tier 232, on the other hand, is
2 implemented at a remote location other than where the server(s) being managed
3 are located (e.g., other than the co-location facility), but from a client computer
4 that is still communicatively coupled to the server(s). The application operations
5 management tier 232 involves managing the software operations of the server(s)
6 and defining sub-boundaries within server clusters. The client can be coupled to
7 the server(s) in any of a variety of manners, such as via the Internet or via a
8 dedicated (e.g., dial-up) connection. The client can be coupled continually to the
9 server(s), or alternatively sporadically (e.g., only when needed for management
10 purposes).

11 The application development tier 234 is implemented on another client
12 computer at a location other than the server(s) (e.g., other than at the co-location
13 facility) and involves development of software components or engines for
14 execution on the server(s). Alternatively, current software on a node 210 at co-
15 location facility 104 could be accessed by a remote client to develop additional
16 software components or engines for the node. Although the client at which
17 application development tier 234 is implemented is typically a different client than
18 that at which application operations management tier 232 is implemented, tiers
19 232 and 234 could be implemented (at least in part) on the same client.

20 Although only three tiers are illustrated in Fig. 4, alternatively the multi-
21 tiered architecture could include different numbers of tiers. For example, the
22 application operations management tier may be separated into two tiers, each
23 having different (or overlapping) responsibilities, resulting in a 4-tiered
24 architecture. The management at these tiers may occur from the same place (e.g.,
25 a single application operations management console may be shared), or

1 alternatively from different places (e.g., two different operations management
2 consoles).

3 Returning to Fig. 3, co-location facility 104 includes a cluster operations
4 management console for each server cluster. In the example of Fig. 3, cluster
5 operations management console 240 corresponds to cluster 212. Cluster
6 operations management console 240 implements cluster operations management
7 tier 230 (Fig. 4) for cluster 212 and is responsible for managing the hardware
8 operations of nodes 210 in cluster 212. Cluster operations management console
9 240 monitors the hardware in cluster 212 and attempts to identify hardware
10 failures. Any of a wide variety of hardware failures can be monitored for, such as
11 processor failures, bus failures, memory failures, etc. Hardware operations can be
12 monitored in any of a variety of manners, such as cluster operations management
13 console 240 sending test messages or control signals to the nodes 210 that require
14 the use of particular hardware in order to respond (no response or an incorrect
15 response indicates failure), having messages or control signals that require the use
16 of particular hardware to generate periodically sent by nodes 210 to cluster
17 operations management console 240 (not receiving such a message or control
18 signal within a specified amount of time indicates failure), etc. Alternatively,
19 cluster operations management console 240 may make no attempt to identify what
20 type of hardware failure has occurred, but rather simply that a failure has occurred.

21 Once a hardware failure is detected, cluster operations management console
22 240 acts to correct the failure. The action taken by cluster operations management
23 console 240 can vary based on the hardware as well as the type of failure, and can
24 vary for different server clusters. The corrective action can be notification of an
25 administrator (e.g., a flashing light, an audio alarm, an electronic mail message,

1 calling a cell phone or pager, etc.), or an attempt to physically correct the problem
2 (e.g., reboot the node, activate another backup node to take its place, etc.).

3 Cluster operations management console 240 also establishes cluster
4 boundaries within co-location facility 104. The cluster boundaries established by
5 console 240 prevent nodes 210 in one cluster (e.g., cluster 212) from
6 communicating with nodes in another cluster (e.g., any node not in cluster 212),
7 while at the same time not interfering with the ability of nodes 210 within a cluster
8 from communicating with other nodes within that cluster. These boundaries
9 provide security for the tenants' data, allowing them to know that their data cannot
10 be communicated to other tenants' nodes 210 at facility 104 even though network
11 connection 216 may be shared by the tenants.

12 In the illustrated example, each cluster of co-location facility 104 includes a
13 dedicated cluster operations management console. Alternatively, a single cluster
14 operations management console may correspond to, and manage hardware
15 operations of, multiple server clusters. According to another alternative, multiple
16 cluster operations management consoles may correspond to, and manage hardware
17 operations of, a single server cluster. Such multiple consoles can manage a single
18 server cluster in a shared manner, or one console may operate as a backup for
19 another console (e.g., providing increased reliability through redundancy, to allow
20 for maintenance, etc.).

21 An application operations management console 242 is also
22 communicatively coupled to co-location facility 104. Application operations
23 management console 242 is located at a location remote from co-location facility
24 104 (that is, not within co-location facility 104), typically being located at the
25 offices of the customer. A different application operations management console

1 242 corresponds to each server cluster of co-location facility 104, although
2 alternatively multiple consoles 242 may correspond to a single server cluster, or a
3 single console 242 may correspond to multiple server clusters. Application
4 operations management console 242 implements application operations
5 management tier 232 (Fig. 4) for cluster 212 and is responsible for managing the
6 software operations of nodes 210 in cluster 212 as well as securing sub-boundaries
7 within cluster 212.

8 Application operations management console 242 monitors the software in
9 cluster 212 and attempts to identify software failures. Any of a wide variety of
10 software failures can be monitored for, such as application processes or threads
11 that are "hung" or otherwise non-responsive, an error in execution of application
12 processes or threads, etc. Software operations can be monitored in any of a variety
13 of manners (similar to the monitoring of hardware operations discussed above),
14 such as application operations management console 242 sending test messages or
15 control signals to particular processes or threads executing on the nodes 210 that
16 require the use of particular routines in order to respond (no response or an
17 incorrect response indicates failure), having messages or control signals that
18 require the use of particular software routines to generate periodically sent by
19 processes or threads executing on nodes 210 to application operations
20 management console 242 (not receiving such a message or control signal within a
21 specified amount of time indicates failure), etc. Alternatively, application
22 operations management console 242 may make no attempt to identify what type of
23 software failure has occurred, but rather simply that a failure has occurred.

24 Once a software failure is detected, application operations management
25 console 242 acts to correct the failure. The action taken by application operations

1 management console 242 can vary based on the hardware as well as the type of
2 failure, and can vary for different server clusters. The corrective action can be
3 notification of an administrator (e.g., a flashing light, an audio alarm, an electronic
4 mail message, calling a cell phone or pager, etc.), or an attempt to correct the
5 problem (e.g., reboot the node, re-load the software component or engine image,
6 terminate and re-execute the process, etc.).

7 Thus, the management of a node 210 is distributed across multiple
8 managers, regardless of the number of other nodes (if any) situated at the same
9 location as the node 210. The multi-tiered management allows the hardware
10 operations management to be separated from the application operations
11 management, allowing two different consoles (each under the control of a different
12 entity) to share the management responsibility for the node.

13 The multi-tiered management architecture can also be used in other
14 situations to manage one or more computers from one or more remote locations,
15 even if the computers are not part of a co-location facility. By way of example, a
16 small business may purchase their own computers, but hire another company to
17 manage the hardware operations of the computers, and possibly yet another
18 company to manage the software operations of the computers.

19 In this example, the small business (the owner of the computers) is a first
20 management tier. The owner then leases the computers to the outsourced
21 hardware operator, which is the second management tier. The hardware operator
22 can manage the hardware operation from a control console, either located locally
23 at the small business along with the computers being managed or alternatively at
24 some remote location, analogous to cluster operations management console 240.

25 The hardware operator then leases the computers to an outsourced software

operator, which is the third management tier. The software operator can manage the software operation from a control console, either located locally at the small business along with the computers being managed or alternatively at some remote location, analogous to application operations management console 242. The software operator then leases the computers back to their owner, so the owner becomes the "user" of the computers, which is the fourth management tier. During normal operation, the computer owner occupies this fourth management tier. However, the computer owner can exercise its first management tier rights to sever one or both of the leases to the software operator and the hardware operator, such as when the computer owner desires to change software or hardware operators.

Fig. 5 is a block diagram illustrating an exemplary node in more detail in accordance with certain embodiments of the invention. Node 248 is an exemplary node managed by other devices (e.g., consoles 240 and 242 of Fig. 3) external to the node. Node 248 can be a node 210 of Fig. 3, or alternatively a node at another location (e.g., a computer in a business or home environment). Node 248 includes a monitor 250, referred to as the "BMonitor", and a plurality of software components or engines 252, and is coupled to (or alternatively incorporates) a mass storage device 262. In the illustrated example, node 248 is a server computer having a processor(s) that supports multiple privilege levels (e.g., rings in an x86 architecture processor). In the illustrated example, these privilege levels are referred to as rings, although alternate implementations using different processor architectures may use different nomenclature. The multiple rings provide a set of prioritized levels that software can execute at, often including 4 levels (Rings 0, 1, 2, and 3). Ring 0 is typically referred to as the most privileged ring. Software

1 processes executing in Ring 0 can typically access more features (e.g.,
2 instructions) than processes executing in less privileged Rings. Furthermore, a
3 processor executing in a particular Ring cannot alter code or data in a higher
4 priority ring. In the illustrated example, BMonitor 250 executes in Ring 0, while
5 engines 252 execute in Ring 1 (or alternatively Rings 2 and/or 3). Thus, the code
6 or data of BMonitor 250 (executing in Ring 0) cannot be altered directly by
7 engines 252 (executing in Ring 1). Rather, any such alterations would have to be
8 made by an engine 252 requesting BMonitor 250 to make the alteration (e.g., by
9 sending a message to BMonitor 250, invoking a function of BMonitor 250, etc.).
10 Implementing BMonitor 250 in Ring 0 protects BMonitor 250 from a rogue or
11 malicious engine 252 that tries to bypass any restrictions imposed by BMonitor
12 250.

13 BMonitor 250 is the fundamental control module of node 248 – it controls
14 (and optionally includes) both the network interface card and the memory
15 manager. By controlling the network interface card (which may be separate from
16 BMonitor 250, or alternatively BMonitor 250 may be incorporated on the network
17 interface card), BMonitor 250 can control data received by and sent by node 248.
18 By controlling the memory manager, BMonitor 250 controls the allocation of
19 memory to engines 252 executing in node 248 and thus can assist in preventing
20 rogue or malicious engines from interfering with the operation of BMonitor 250.

21 Although various aspects of node 248 may be under control of BMonitor
22 250 (e.g., the network interface card), BMonitor 250 still makes at least part of
23 such functionality available to engines 252 executing on the node 248. BMonitor
24 250 provides an interface (e.g., via controller 254 discussed in more detail below)
25 via which engines 252 can request access to the functionality, such as to send data

1 out to another node 248 or to the Internet. These requests can take any of a variety
2 of forms, such as sending messages, calling a function, etc.

3 BMonitor 250 includes controller 254, network interface 256, one or more
4 filters 258, and a Distributed Host Control Protocol (DHCP) module 260.
5 Network interface 256 provides the interface between node 248 and the network
6 (e.g., network connections 126 of Fig. 3) via the internal transport medium 211 of
7 co-location facility 104. Filters 258 identify other nodes 248 (and/or other sources
8 or targets (e.g., coupled to Internet 108 of Fig. 1) that data can (or alternatively
9 cannot) be sent to and/or received from. The nodes or other sources/targets can be
10 identified in any of a wide variety of manners, such as by network address (e.g.,
11 Internet Protocol (IP) address), some other globally unique identifier, a locally
12 unique identifier (e.g., a numbering scheme proprietary or local to co-location
13 facility 104), etc.

14 Filters 258 can fully restrict access to a node (e.g., no data can be received
15 from or sent to the node), or partially restrict access to a node. Partial access
16 restriction can take different forms. For example, a node may be restricted so that
17 data can be received from the node but not sent to the node (or vice versa). By
18 way of another example, a node may be restricted so that only certain types of data
19 (e.g., communications in accordance with certain protocols, such as HTTP) can be
20 received from and/or sent to the node. Filtering based on particular types of data
21 can be implemented in different manners, such as by communicating data in
22 packets with header information that indicate the type of data included in the
23 packet.

24 Filters 258 can be added by application operations management console
25 242 or cluster operations management console 240. In the illustrated example,

filters added by cluster operations management console 240 (to establish cluster boundaries) restrict full access to nodes (e.g., any access to another node can be prevented) whereas filters added by application operations management console 242 (to establish sub-boundaries within a cluster) can restrict either full access to nodes or partial access.

Controller 254 also imposes some restrictions on what filters can be added to filters 258. In the illustrated example, controller 254 allows cluster operations management console 240 to add any filters it desires (which will define the boundaries of the cluster). However, controller 254 restricts application operations management console 242 to adding only filters that are at least as restrictive as those added by console 240. If console 242 attempts to add a filter that is less restrictive than those added by console 240 (in which case the sub-boundary may extend beyond the cluster boundaries), controller 254 refuses to add the filter (or alternatively may modify the filter so that it is not less restrictive). By imposing such a restriction, controller 254 can ensure that the sub-boundaries established at the application operations management level do not extend beyond the cluster boundaries established at the cluster operations management level.

Controller 254, using one or more filters 258, operates to restrict data packets sent from node 248 and/or received by node 248. All data intended for an engine 252, or sent by an engine 252, to another node, is passed through network interface 256 and filters 258. Controller 254 applies the filters 258 to the data, comparing the target of the data (e.g., typically identified in a header portion of a packet including the data) to acceptable (and/or restricted) nodes (and/or network addresses) identified in filters 258. If filters 258 indicate that the target of the data is acceptable, then controller 254 allows the data to pass through to the target

1 (either into node 248 or out from node 248). However, if filters 258 indicate that
2 the target of the data is not acceptable, then controller 254 prevents the data from
3 passing through to the target. Controller 254 may return an indication to the
4 source of the data that the data cannot be passed to the target, or may simply
5 ignore or discard the data.

6 The application of filters 258 to the data by controller 254 allows the
7 boundary restrictions of a server cluster to be imposed. Filters 258 can be
8 programmed (e.g., by application operations management console 242 of Fig. 3)
9 with the node addresses of all the nodes within the server cluster (e.g., cluster
10 212). Controller 254 then prevents data received from any node not within the
11 server cluster from being passed through to an engine 252, and similarly prevents
12 any data being sent to a node other than one within the server cluster from being
13 sent. Similarly, data received from Internet 108 (Fig. 1) can identify a target node
14 210 (e.g., by IP address), so that controller 254 of any node other than the target
15 node will prevent the data from being passed through to an engine 252.

16 DHCP module 260 implements the Distributed Host Control Protocol,
17 allowing BMonitor 250 (and thus node 210) to obtain an IP address from a DHCP
18 server (e.g., cluster operations management console 240 of Fig. 3). During an
19 initialization process for node 210, DHCP module 260 requests an IP address from
20 the DHCP server, which in turn provides the IP address to module 260.
21 Additional information regarding DHCP is available from Microsoft Corporation
22 of Redmond, Washington.

23 Software engines 252 include any of a wide variety of conventional
24 software components. Examples of engines 252 include an operating system (e.g.,
25 Windows NT®), a load balancing server component (e.g., to balance the

1 processing load of multiple nodes 248), a caching server component (e.g., to cache
2 data and/or instructions from another node 248 or received via the Internet), a
3 storage manager component (e.g., to manage storage of data from another node
4 248 or received via the Internet), etc. In one implementation, each of the engines
5 252 is a protocol-based engine, communicating with BMonitor 250 and other
6 engines 252 via messages and/or function calls without requiring the engines 252
7 and BMonitor 250 to be written using the same programming language.

8 Controller 254 is further responsible for controlling the execution of
9 engines 252. This control can take different forms, including beginning execution
10 of an engine 252, terminating execution of an engine 252, re-loading an image of
11 an engine 252 from a storage device, debugging execution of an engine 252, etc.
12 Controller 254 receives instructions from application operations management
13 console 242 of Fig. 3 regarding which of these control actions to take and when to
14 take them. Thus, the control of engines 252 is actually managed by the remote
15 application operations management console 242, not locally at co-location facility
16 104. Controller 254 also provides an interface via which application operations
17 management console 242 can identify filters to add (and/or remove) from filter set
18 258.

19 Controller 254 also includes an interface via which cluster operations
20 management console 240 of Fig. 3 can communicate commands to controller 254.
21 Different types of hardware operation oriented commands can be communicated to
22 controller 254 by cluster operations management console 240, such as re-booting
23 the node, shutting down the node, placing the node in a low-power state (e.g., in a
24 suspend or standby state), changing cluster boundaries, changing encryption keys,
25 etc.

1 Controller 254 further provides encryption support for BMonitor 250,
2 allowing data to be stored securely on mass storage device 262 (e.g., a magnetic
3 disk, an optical disk, etc.) and secure communications to occur between node 248
4 and an operations management console (e.g., console 240 or 242 of Fig. 3).
5 Controller 254 maintains multiple encryption keys, including: one for the landlord
6 (referred to as the "landlord key") which accesses node 248 from cluster
7 operations management console 240, one for the lessee of node 248 (referred to as
8 the "tenant key") which accesses node 248 from application operations
9 management console 242, and keys that BMonitor 250 uses to securely store data
10 on mass storage device 262 (referred to as the "disk key").

11 BMonitor 250 makes use of public key cryptography to provide secure
12 communications between node 248 and the management consoles (e.g., consoles
13 240 and 242). Public key cryptography is based on a key pair, including both a
14 public key and a private key, and an encryption algorithm. The encryption
15 algorithm can encrypt data based on the public key such that it cannot be
16 decrypted efficiently without the private key. Thus, communications from the
17 public-key holder can be encrypted using the public key, allowing only the
18 private-key holder to decrypt the communications. Any of a variety of public key
19 cryptography techniques may be used, such as the well-known RSA (Rivest,
20 Shamir, and Adelman) encryption technique. For a basic introduction of
21 cryptography, the reader is directed to a text written by Bruce Schneier and
22 entitled "Applied Cryptography: Protocols, Algorithms, and Source Code in C,"
23 published by John Wiley & Sons with copyright 1994 (or second edition with
24 copyright 1996).

1 BMonitor 250 is initialized to include a public/private key pair for both the
2 landlord and the tenant. These key pairs can be generated by BMonitor 250, or
3 alternatively by some other component and stored within BMonitor 250 (with that
4 other component being trusted to destroy its knowledge of the key pair). As used
5 herein, U refers to a public key and R refers to a private key. The public/private
6 key pair 264 for the landlord is referred to as (U_L, R_L) , and the public/private key
7 pair 266 for the tenant is referred to as (U_T, R_T) . BMonitor 250 makes the public
8 keys U_L and U_T available to the landlord, but keeps the private keys R_L and R_T
9 secret. In the illustrated example, BMonitor 250 never divulges the private keys
10 R_L and R_T , so both the landlord and the tenant can be assured that no entity other
11 than the BMonitor 250 can decrypt information that they encrypt using their public
12 keys (e.g., via cluster operations management console 240 and application
13 operations management console 242 of Fig. 3, respectively).

14 Once the landlord has the public keys U_L and U_T , the landlord can assign
15 node 210 to a particular tenant, giving that tenant the public key U_T . Use of the
16 public key U_T allows the tenant to encrypt communications to BMonitor 250 that
17 only BMonitor 250 can decrypt (using the private key R_T). Although not required,
18 a prudent initial step for the tenant is to request that BMonitor 250 generate a new
19 public/private key pair (U_T, R_T) . In response to such a request, a key generator 268
20 of BMonitor 250 generates a new public/private key pair in any of a variety of
21 well-known manners, stores the new key pair as key pair 266, and returns the new
22 public key U_T to the tenant. By generating a new key pair, the tenant is assured
23 that no other entity, including the landlord, is aware of the tenant public key U_T .
24 Additionally, the tenant may also have new key pairs generated at subsequent
25 times.

1 BMonitor 250 enforces restrictions on what entities can request new
2 public/private key pairs. The tenant is able to request new tenant public/private
3 key pairs, but is not able to request new landlord public/private key pairs. The
4 landlord, however, can request new landlord public/private key pairs as well as
5 new tenant public/private key pairs. Whenever a request for a new public/private
6 key pair is received, controller 254 verifies the identity of the requestor as the
7 tenant or landlord (e.g., based on a remote log-in procedure, password verification,
8 manner in which the requestor is communicating with or is coupled to node 248,
9 etc.) before generating the new key pair.

10 In order to ensure bi-directional communication security between BMonitor
11 250 and the landlord and tenant control devices (e.g., operations management
12 consoles 240 and 242, respectively), the landlord and tenant control devices may
13 also generate (or otherwise be assigned) public/private key pairs. In this situation,
14 consoles 240 and 242 can communicate their respective public keys to BMonitors
15 250 of nodes 248 they desire (or expect to desire) to communicate with securely.
16 Once the public key of a console is known by a BMonitor 250, the BMonitor 250
17 can encrypt communications to that console using its public key, thereby
18 preventing any other device except the console having the private key from
19 reading the communication.

20 BMonitor 250 also maintains a disk key 270, which is generated based on
21 one or more symmetric keys 272 and 274 (symmetric keys refer to secret keys
22 used in secret key cryptography). Disk key 270, also a symmetric key, is used by
23 BMonitor 250 to store information in mass storage device 262. BMonitor 250
24 keeps disk key 270 secure, using it only to encrypt data node 248 stores on mass
25 storage device 262 and decrypt data node 248 retrieves from mass storage device

1 262 (thus there is no need for any other entities, including the landlord and tenant,
2 to have knowledge of disk key 270). Alternatively, the landlord or tenant may be
3 informed of disk key 270, or another key on which disk key 270 is based.

4 Use of disk key 270 ensures that data stored on mass storage device 262
5 can only be decrypted by the node 248 that encrypted it, and not any other node or
6 device. Thus, for example, if mass storage device 262 were to be removed and
7 attempts made to read the data on device 262, such attempts would be
8 unsuccessful. BMonitor 250 uses disk key 270 to encrypt data to be stored on
9 mass storage device 262 regardless of the source of the data. For example, the
10 data may come from a client device (e.g., client 102 of Fig. 1) used by a customer
11 of the tenant, from an operations management console (e.g., console 242 of Fig.
12 3), etc.

13 Disk key 270 is generated based on symmetric keys 272 and 274. As used
14 herein, K refers to a symmetric key, so K_L refers to a landlord symmetric key (key
15 272) and K_T refers to a tenant symmetric key (key 274). The individual keys 272
16 and 274 can be generated in any of a wide variety of conventional manners (e.g.,
17 based on a random number generator). Disk key 270 is either the K_L key alone, or
18 alternatively is a combination of the K_L and K_T keys. In situations where the node
19 210 is not currently leased to a tenant, or in which the tenant has not established a
20 K_T key, then controller 254 maintains the K_L key as disk key 270. However, in
21 situations where the node 248 is leased to a tenant that establishes a K_T key, then
22 disk key 270 is a combination of the K_L and K_T keys. The K_L and K_T keys can be
23 combined in a variety of different manners, and in one implementation are
24 combined by using one of the keys to encrypt the other key, with the resultant
25 encrypted key being disk key 270. Thus, the data stored on mass storage device

1 262 is always encrypted, even if the tenant does not establish a symmetric key K_T .
2 Additionally, in situations where the landlord and tenant are aware of their
3 respective keys K_L and K_T , then the combination of the keys results in a key that
4 can be used to encrypt the data so that neither the landlord nor the tenant can
5 decrypt it individually.

6 In the illustrated example, a node 248 does not initially have symmetric
7 keys K_L and K_T . When the landlord initializes the node 248, it requests a new key
8 K_L (e.g., via cluster operations management console 240 of Fig. 3), in response to
9 which key generator 268 generates a new key and controller 254 maintains the
10 newly generated key as key 272. Similarly, when a tenant initially leases a node
11 248 there is not yet a tenant symmetric key K_T for node 248. The tenant can
12 communicate a request for a new key K_T (e.g., via application operations
13 management console 242 of Fig. 3), in response to which key generator 268
14 generates a new key and controller 254 maintains the newly generated key as key
15 274. Additionally, each time a new key K_T or K_L is generated, then controller 254
16 generates a new disk key 270.

17 Although only a landlord and tenant key (K_L and K_T) are illustrated in
18 Fig. 5, alternatively additional symmetric keys (e.g., from a sub-tenant, a sub-
19 landlord, etc.) may be combined to generate disk key 270. For example, if there
20 are three symmetric keys, then they can be combined by encrypting a first of the
21 keys with a second of the keys, and then encrypting the result with the third of the
22 keys to generate disk key 270. Additional symmetric keys may be used, for
23 example, for a sub-tenant(s).

24 The landlord can also request new public/private key pairs from BMonitor
25 250, either tenant key pairs or landlord key pairs. Requesting new key pairs can

allow, for example, the landlord to re-assign a node 248 from one tenant to another. By way of example, if a tenant no longer desires the node 248 (or does not make required lease payments for the node), then the landlord can communicate with BMonitor 250 (e.g., via console 240 of Fig. 3) to change the public/private key pairs of the tenant (thereby prohibiting any communications from the tenant from being decrypted by the BMonitor 250 because the tenant does not have the new key). Additionally, the landlord may also request a new public/private key pair for the landlord – this may be done at particular intervals or simply whenever the landlord desires a new key (e.g., for safety concerns).

In one implementation, BMonitor 250 discards both the disk key 270 and the landlord symmetric key K_L , and generates a new key K_L (and a new disk key 270) each time it generates a new landlord private key R_L . By replacing the key K_L and disk key 270 (and keeping no record of the old keys), the landlord can ensure that once it changes its key, any tenant data previously stored at the node 210 cannot be accessed. Thus, care should be taken by the landlord to generate a new public/private key pair only when the landlord wants to prevent the tenant from accessing the data previously stored at node 248.

Additionally, BMonitor 250 may also replace both the disk key 270 and the tenant symmetric key K_T , with a newly generated key K_T (and a new disk key 270) each time it generates a new tenant private key R_T . This allows the tenant to increase the security of the data being stored at the node 248 because it can change how that data is encrypted as it desires. However, as BMonitor 250 discards the previous key K_T and disk key 270, care should be exercised by the tenant to request a new tenant private key R_T only when the data previously stored at node 210 is no longer needed (e.g., has been backed up elsewhere).

1 It should be noted that different nodes 248 will typically have different keys
2 (keys 264, 266, and 270). Alternatively, attempts may be made to have multiple
3 nodes use the same key (e.g., key 270). However, in such situations care should
4 be taken to ensure that any communication of the keys (e.g., between nodes 248)
5 is done in a secure manner so that the security is not compromised. For example,
6 additional public/private key pairs may be used by BMonitors 250 of two nodes
7 248 to securely communicate information between one another.

8 A leased hardware environment having guaranteed and enforced rights can
9 thus be established. Landlords can lease nodes to multiple different tenants and
10 establish boundaries that prevent nodes leased by different tenants from
11 communicating with one another. Tenants can be assured that nodes they lease are
12 accessible for management only to them, not to others, and that data is stored at
13 the nodes securely so that no one else can access it (even if the tenant leaves or
14 reduces its hardware usages). Furthermore, landlords and tenants are both assured
15 that the landlord can move equipment, change which nodes are assigned to
16 individuals, remove hardware (e.g., mass storage devices), etc. without
17 compromising the secure storage of data by any of the tenants.

18 Fig. 6 is a flowchart illustrating an exemplary process for encryption key
19 generation and distribution in accordance with certain embodiments of the
20 invention. Initially, the computer (e.g., a node 248 of Fig. 5) identifies
21 public/private key pairs for both the landlord and the tenant (act 280). This
22 identification can be accessing previously generated key pairs, or alternatively
23 generating a new key pair by the computer itself. The computer keeps both the
24 landlord private key from the landlord key pair and the tenant private key from the
25 tenant key pair secret, but forwards the landlord public key from the landlord key

1 pair and the tenant public key from the tenant key pair to the landlord (act 282). In
2 the illustrated example, the landlord is represented by cluster operations
3 management console 240 of Fig. 3, although alternatively other devices or entities
4 could represent the landlord.

5 The landlord then forwards the tenant public key to the tenant (act 284). In
6 the illustrated example, the tenant is represented by application operations
7 management console 242 of Fig. 3, although alternatively other devices or entities
8 could represent the tenant. The tenant then communicates with the computer to
9 generate a new tenant key pair (act 286). The computer keeps the tenant private
10 key from the new key pair secret and forwards the tenant public key from the new
11 key pair to the tenant (act 288). The tenant is then able to communicate secure
12 messages (e.g., data, instructions, requests, etc.) to the computer using the new
13 tenant public key (act 290), while the landlord is able to communicate secure
14 messages to the computer using the landlord public key (act 292).

15 Fig. 7 is a flowchart illustrating an exemplary process for the operation of a
16 cluster operations management console in accordance with certain embodiments
17 of the invention. The process of Fig. 7 is implemented by a cluster operations
18 management console at a co-location facility, and may be performed in software.

19 Initially, the cluster operations management console configures the nodes in
20 the server cluster with the boundaries (if any) of the server cluster (act 300). This
21 configuration is accomplished by the cluster operations management console
22 communicating filters to the nodes in the server cluster(s).

23 Hardware operations within a server cluster are then continually monitored
24 for a hardware failure (acts 302 and 304). Once a hardware failure is detected,
25 corrective action is taken (act 306) and monitoring of the hardware operation

continues. Any of a wide variety of corrective action can be taken, as discussed above. Note that, based on the corrective action (or at other times), the nodes may be re-configured with new cluster boundaries (act 300).

Fig. 8 is a flowchart illustrating an exemplary process for the operation of an application operations management console in accordance with certain embodiments of the invention. The process of Fig. 8 is implemented by an application operations management console located remotely from the co-location facility, and may be performed in software.

Initially, the application operations management console configures the nodes in the server cluster with sub-boundaries (if any) of the server cluster (act 320). This configuration is accomplished by the application operations management console communicating filters to the nodes in the server cluster.

Software operations within the server cluster are then continually monitored until a software failure is detected (acts 322 and 324). This software failure could be failure of a particular software engine (e.g., the engine fails, but the other engines are still running), or alternatively failure of the entire node (e.g., the entire node is hung). Once a software failure is detected, corrective action is taken (act 326) and monitoring of the software operation continues. Any of a wide variety of corrective action can be taken, as discussed above. Note that, based on the corrective action (or at any other time during operation), the server computer may be re-configured with new sub-boundaries (act 320).

Conclusion

Although the description above uses language that is specific to structural features and/or methodological acts, it is to be understood that the invention

1 defined in the appended claims is not limited to the specific features or acts
2 described. Rather, the specific features and acts are disclosed as exemplary forms
3 of implementing the invention.

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

1 **CLAIMS**

- 2
- 3 1. A multi-tiered management architecture comprising:
4 an application development tier at which applications are developed for
5 execution on one or more computers;
6 an application operations tier at which execution of the applications is
7 managed; and
8 a cluster operations tier to manage the operation of the computers without
9 concern for what applications are executing on the one or more computers.
- 10
- 11 2. A management architecture as recited in claim 1, wherein the cluster
12 operations tier is responsible for securing a computer cluster boundary to prevent a
13 plurality of other computers that are not part of the computer cluster from
14 accessing the one or more computers in the computer cluster.
- 15
- 16 3. A management architecture as recited in claim 1, wherein the
17 application operations tier is responsible for securing sub-boundaries within the
18 computer cluster boundary to restrict communication between computers within
19 the computer cluster.
- 20
- 21 4. A management architecture as recited in claim 1, wherein the
22 application operations tier is implemented at an application operations
23 management console at a location remote from the one or more computers.
- 24
- 25

1 **5.** A management architecture as recited in claim 1, wherein the cluster
2 operations tier is implemented at a cluster operations management console located
3 at the same location as the one or more computers.

4

5 **6.** A management architecture as recited in claim 1, wherein the
6 application operations tier monitors execution of application processes on the one
7 or more computers and detects failures of the application processes.

8

9 **7.** A management architecture as recited in claim 1, wherein the
10 application operations tier takes corrective action in response to a software failure
11 on one of the computers.

12

13 **8.** A management architecture as recited in claim 7, wherein the
14 corrective action comprises re-booting the computer.

15

16 **9.** A management architecture as recited in claim 7, wherein the
17 corrective action comprises notifying an administrator of the failure.

18

19 **10.** A management architecture as recited in claim 1, wherein the cluster
20 operations tier monitors hardware operation of the one or more computers and
21 detects failures of the hardware.

1 **11.** A management architecture as recited in claim 1, wherein the cluster
2 operations tier takes corrective action in response to a hardware failure of one of
3 the computers.

4

5 **12.** A management architecture as recited in claim 11, wherein the
6 corrective action comprises re-booting the computer.

7

8 **13.** A management architecture as recited in claim 11, wherein the
9 corrective action comprises notifying a co-location facility administrator.

10

11 **14.** A management architecture as recited in claim 11, wherein the one
12 or more computers are situated in one or more clusters at a co-location facility.

13

14 **15.** A co-location facility system comprising:
15 a plurality of node clusters each corresponding to a different customer; and
16 a cluster operations management console corresponding to at least one of
17 the node clusters and configured to manage hardware operations of the at least one
18 node cluster.

19

20 **16.** A system as recited in claim 15, further comprising a different
21 cluster operations management console corresponding to each of the plurality of
22 node clusters.

1 **17.** A system as recited in claim 15, wherein each of the plurality of
2 node clusters includes, as its nodes, a plurality of server computers.
3

4 **18.** A system as recited in claim 15, wherein the hardware operations
5 include one or more of: mass storage device operation, memory device operation,
6 and network interface operation, and processor operation.
7

8 **19.** A system as recited in claim 15, wherein each of the plurality of
9 node clusters includes a plurality of nodes configured to receive node control
10 commands from an application operations management console located remotely
11 from the co-location facility.
12

13 **20.** A system as recited in claim 19, wherein each node in each node
14 cluster is configured with a private key that allows the node to decrypt
15 communications that are received, in a form encrypted using a public key, from
16 the application operations management console associated with the customer that
17 corresponds to the node cluster.
18

19 **21.** A system as recited in claim 15, further comprising a data transport
20 medium coupled to each node in the plurality of clusters via which each node can
21 access an external network.
22

23 **22.** A system as recited in claim 15, wherein the external network
24 comprises the Internet.
25

1 **23.** A system as recited in claim 15, wherein each node in each node
2 cluster is configured with the boundary of the node cluster.

3

4 **24.** A system as recited in claim 15, wherein each node in each node
5 cluster is configured with a private key that allows the node to decrypt
6 communications that are received, in a form encrypted using a public key, from
7 the cluster operations management console.

8

9 **25.** A system as recited in claim 15, wherein one or more of the nodes in
10 a node cluster are leased by the customer from an operator of the co-location
11 facility.

12

13 **26.** A method comprising:
14 monitoring, at a co-location facility, hardware operations of a cluster of
15 computers located at the co-location facility;
16 detecting a hardware failure in one of the computers in the cluster; and
17 performing an act, in response to detecting the hardware failure, to correct
18 the hardware failure.

19

20 **27.** A method as recited in claim 26, wherein the cluster of computers is
21 one of a plurality of clusters of computers located at the co-location facility.

22

23 **28.** A method as recited in claim 26, wherein the act comprises
24 notifying a co-location facility administrator of the failure.

1 **29.** A method as recited in claim 26, wherein the act comprises resetting
2 the computer that includes the hardware that failed.
3

4 **30.** A method as recited in claim 26, wherein the hardware operation
5 includes one or more of: mass storage device operation, memory device
6 operation, and network interface operation, and processor operation.
7

8 **31.** A method as recited in claim 26, further comprising configuring
9 each computer in the cluster to impose boundaries preventing a plurality of other
10 computers that are not part of the cluster from accessing the one or more
11 computers in the cluster.
12

13 **32.** One or more computer-readable memories containing a computer
14 program that is executable by a processor to perform the method recited in claim
15 26.
16

17 **33.** A method comprising:
18 monitoring, from a location remote from a co-location facility, software
19 operations of a cluster of computers located at the co-location facility;
20 detecting a software failure in one of the computers in the cluster; and
21 performing an act, in response to detecting the software failure, to correct
22 the hardware failure.
23
24
25

1 **34.** A method as recited in claim 33, wherein the cluster of computers is
2 one of a plurality of clusters of computers located at a co-location facility.
3

4 **35.** A method as recited in claim 33, wherein the act comprises
5 notifying an administrator of the failure.
6

7 **36.** A method as recited in claim 33, wherein the act comprises resetting
8 the computer that executes the software that failed.
9

10 **37.** A method as recited in claim 33, further comprising configuring one
11 or more computers in the cluster to impose sub-boundaries preventing a first one
12 or more computers within the cluster from accessing a second one or more
13 computers within the cluster.
14

15 **38.** A method as recited in claim 33, further comprising managing
16 loading of a software component on one of the computers in the cluster.
17

18 **39.** A method as recited in claim 33, wherein the software failure
19 comprises one or more of: a hung application process, a hung thread, and an error
20 in execution of an application process.
21

22 **40.** A computer, located remotely from the cluster of computers, to
23 implement the method of claim 33.
24

1 **41.** A method as recited in claim 33, wherein the monitoring, detecting,
2 and performing are implemented in a remote computer, and further comprising
3 using public key cryptography to securely communicate between the remote
4 computer and each computer in the cluster of computers.

5

6 **42.** One or more computer-readable memories containing a computer
7 program that is executable by a processor to perform the method recited in claim
8 33.

9

10 **43.** One or more computer-readable media having stored thereon a
11 computer program that, when executed by one or more processors, causes the one
12 or more processors to perform acts including:

13 monitoring, from a location remote from a co-location facility, software
14 operations of a cluster of computers located at the co-location facility; and

15 taking corrective action in response to a failure in operation of software
16 executing on one of the computers in the cluster.

17

18 **44.** One or more computer-readable media as recited in claim 43,
19 wherein the corrective action comprises notifying an administrator of the failure.

20

21 **45.** One or more computer-readable media as recited in claim 43,
22 wherein the corrective action comprises resetting the computer that executes the
23 software that failed.

1 **46.** One or more computer-readable media as recited in claim 43,
2 wherein the computer program, when executed, further causes the one or more
3 processors to perform acts including configuring each computer in the cluster to
4 impose boundaries preventing a plurality of other computers that are not part of
5 the cluster from accessing the one or more computers in the cluster.

6

7 **47.** One or more computer-readable media as recited in claim 43,
8 wherein the failure in operation of the software comprises one or more of: a hung
9 application process, a hung thread, and an error in execution of an application
10 process.

11

12 **48.** A method comprising:
13 selling, to a customer, rights to a plurality of computers to be located at a
14 facility; and

15 enforcing a multiple-tiered management scheme on the plurality of
16 computers in which hardware operation of the plurality of computers is managed
17 locally at the facility and software operation of the plurality of computers is
18 managed from a location remote from the facility.

19

20 **49.** A method as recited in claim 48, wherein the facility comprises a
21 co-location facility.

22

23 **50.** A method as recited in claim 48, wherein the selling comprises
24 licensing at least one of the plurality of computers to the customer.

1 **51.** A method as recited in claim 48, wherein the selling comprises
2 selling at least one of the plurality of computers to the customer.

3
4 **52.** A method comprising:

5 allowing a tenant to which a cluster of one or more computers at a facility
6 have been leased to communicate with the one or more computers; and
7 implementing cluster boundaries at the facility to prevent computers within
8 the cluster from communicating with computers in another cluster.

9
10 **53.** A method as recited in claim 52, wherein the facility comprises a
11 co-location facility.

12
13 **54.** A method as recited in claim 52, wherein the allowing comprises
14 establishing secure communications channels between the one or more computers
15 and a corresponding tenant operations management console.

16
17 **55.** A method as recited in claim 54, wherein:
18 the cluster boundaries are implemented at a first tier of a multi-tiered
19 management architecture; and
20 allowing the tenant operations management console, implemented in a
21 second tier of the multi-tiered management architecture, to establish sub-
22 boundaries within the cluster.

1 **56.** A method as recited in claim 52, further comprising:

2 performing at least some management of the computers via a landlord
3 operations management console; and

4 establishing secure communications channels between each of the plurality
5 of computers and the landlord operations management console.

6

7 **57.** A method as recited in claim 56, wherein the landlord operations
8 management console is located at the facility.

9

10 **58.** A method as recited in claim 52, wherein the cluster includes one or
11 more additional computers that have not been leased to the tenant.

12

13 **59.** A method comprising:

14 separating a plurality of computers at a co-location facility into a plurality
15 of clusters;

16 leasing the clusters to a plurality of tenants; and

17 allowing secure communications channels to be established between the
18 computers in the cluster leased to the tenant and an operations management
19 console of the tenant.

20

21 **60.** A method as recited in claim 59, further comprising:

22 performing at least some management of the computers via a landlord
23 operations management console; and

1 allowing secure communications channels to be established between the
2 computers in the cluster leased to the tenant and the landlord operations
3 management console.

4

5 **61.** A method as recited in claim 59, further comprising implementing
6 cluster boundaries to prevent computers within a cluster from communicating with
7 computers in another cluster.

8

9 **62.** A method as recited in claim 59, further comprising implementing
10 cluster sub-boundaries to restrict the ability of computers within a cluster to
11 communicate with other computers within the cluster.

12

13 **63.** A method comprising:
14 generating, at a computer, a landlord key pair and a tenant key pair, each
15 key pair including a private key and a public key, the landlord key pair being used
16 to establish secure communication between the computer and a landlord device,
17 and the tenant key pair being used to establish secure communication between the
18 computer and a tenant device;

19 keeping the landlord private key and the tenant private key secure at the
20 computer without disclosing the keys to any other device;

21 forwarding the landlord public key and the tenant public key to the landlord
22 device; and

23 forwarding the tenant public key to the tenant device.

1 **64.** A method as recited in claim 63, further comprising generating a
2 storage key to encrypt data to be stored on a mass storage device.

3

4 **65.** A method as recited in claim 64, further comprising discarding the
5 current storage key each time the tenant private key is changed, and generating a
6 new storage key.

7

8 **66.** A method as recited in claim 64, wherein the generating the storage
9 key comprises combining a landlord symmetric key and a tenant symmetric key to
10 generate the storage key.

11

12 **67.** A method as recited in claim 63, further comprising forwarding the
13 tenant key to the tenant device via the landlord device.

14

15 **68.** A method as recited in claim 63, wherein the landlord device
16 comprises a cluster operations management console.

17

18 **69.** A method as recited in claim 63, wherein the tenant device
19 comprises an application operations management console.

20

21 **70.** A method comprising:
22 maintaining, at a computer, a storage key to encrypt data to be stored on a
23 mass storage device;
24 using, as the storage key, only a landlord key if no tenant key has been
25 generated at the computer; and

1 if a tenant key has been generated at the computer, then combining the
2 landlord key and the tenant key to generate the storage key.

3

4 **71.** A method as recited in claim 70, further comprising combining the
5 landlord key and the tenant key by using the landlord key to encrypt the tenant
6 key.

7

8 **72.** A method as recited in claim 70, further comprising combining the
9 landlord key and the tenant key by using the tenant key to encrypt the landlord
10 key.

11

12 **73.** A multi-tiered computer management architecture comprising:
13 a first tier corresponding to an owner of a computer;
14 a second tier corresponding to a hardware operator that is to manage
15 hardware operations of the computer;
16 a third tier corresponding to a software operator that is to manage software
17 operations of the computer; and
18 a fourth tier corresponding to the owner, wherein the owner operates in the
19 fourth tier except when revoking the rights of the hardware operator or software
20 operator.

21

22 **74.** An architecture as recited in claim 73, wherein the second tier
23 management is implemented at a management console at a location remote from
24 the computer.

1 **75.** An architecture as recited in claim 73, wherein the third tier
2 management is implemented at a management console at a location remote from
3 the computer.

4

5 **76.** An architecture as recited in claim 73, further comprising using a
6 plurality key pairs, each key pair including a private key and a public key, to
7 securely communicate between the computer and a management device
8 corresponding to the hardware operator, as well as between the computer and a
9 management device corresponding to the software operator.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 **ABSTRACT**

2 A multi-tiered server management architecture is employed including an
3 application development tier, an application operations tier, and a cluster
4 operations tier. In the application development tier, applications are developed for
5 execution on one or more server computers. In the application operations tier,
6 execution of the applications is managed and sub-boundaries within a cluster of
7 servers can be established. In the cluster operations tier, operation of the server
8 computers is managed without concern for what applications are executing on the
9 one or more server computers and boundaries between clusters of servers can be
10 established. The multi-tiered server management architecture can also be
11 employed in co-location facilities where clusters of servers are leased to tenants,
12 with the tenants implementing the application operations tier and the facility
13 owner (or operator) implementing the cluster operations tier.

14
15
16
17
18
19
20
21
22
23
24
25

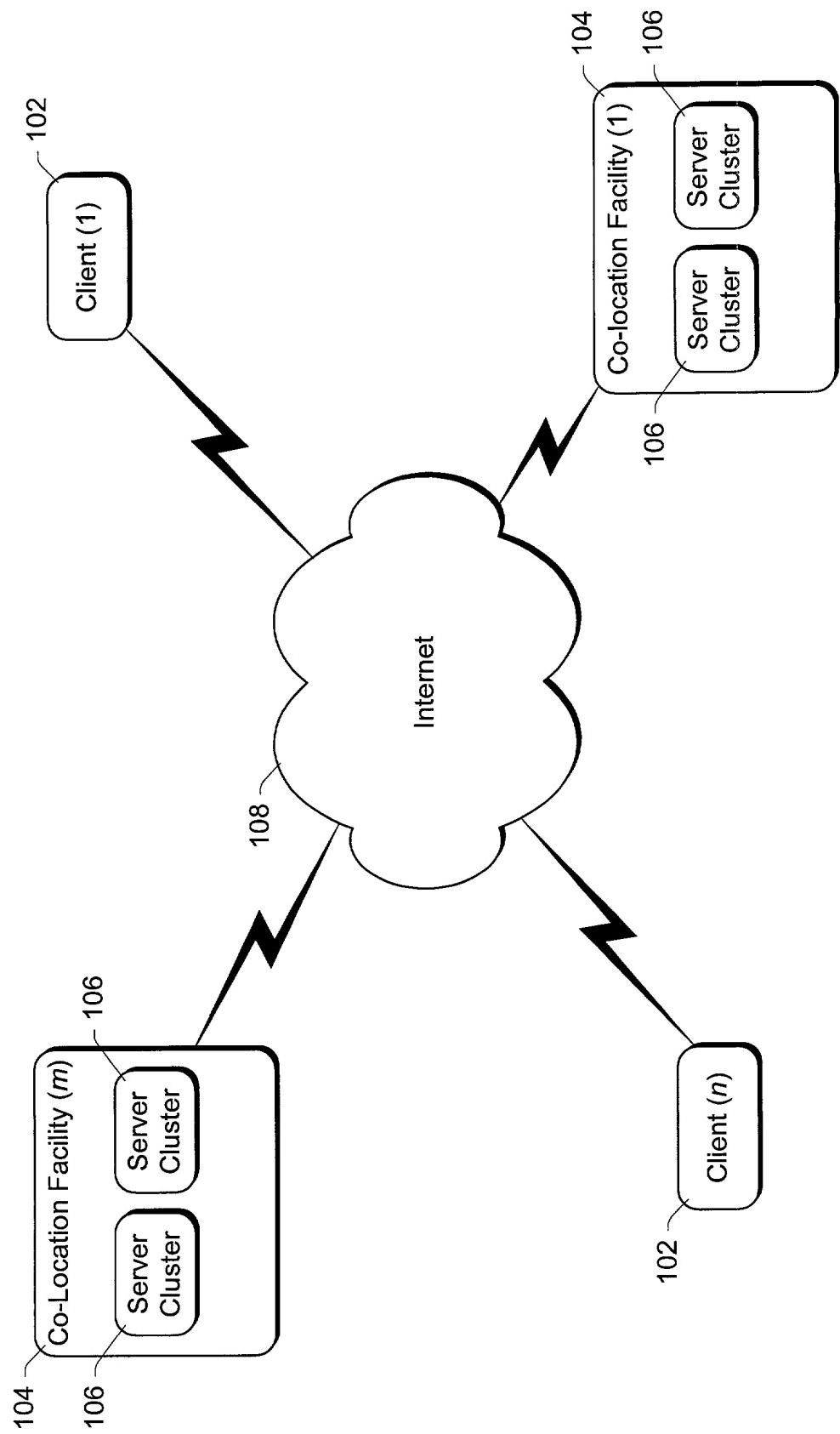
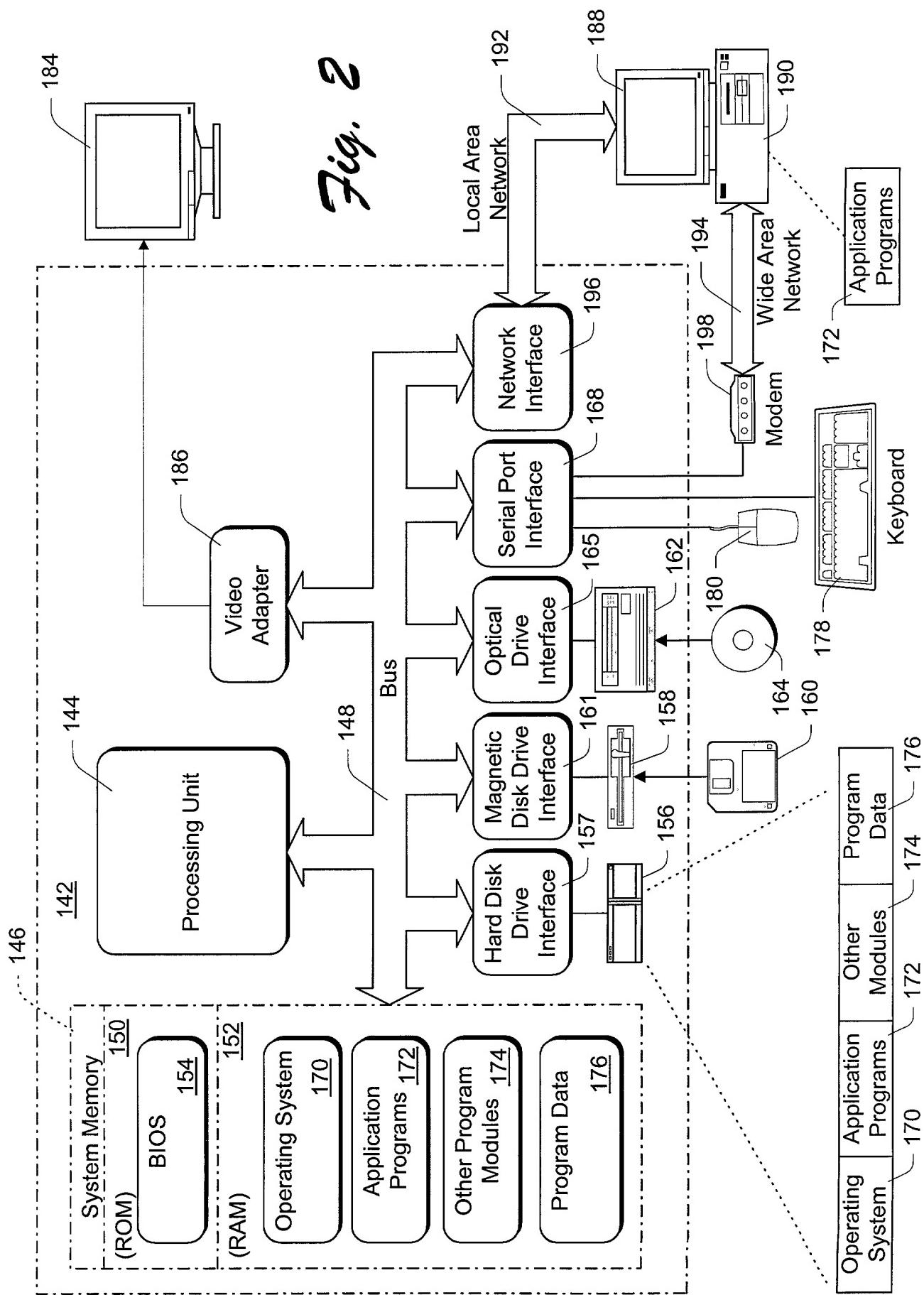


Fig. 1



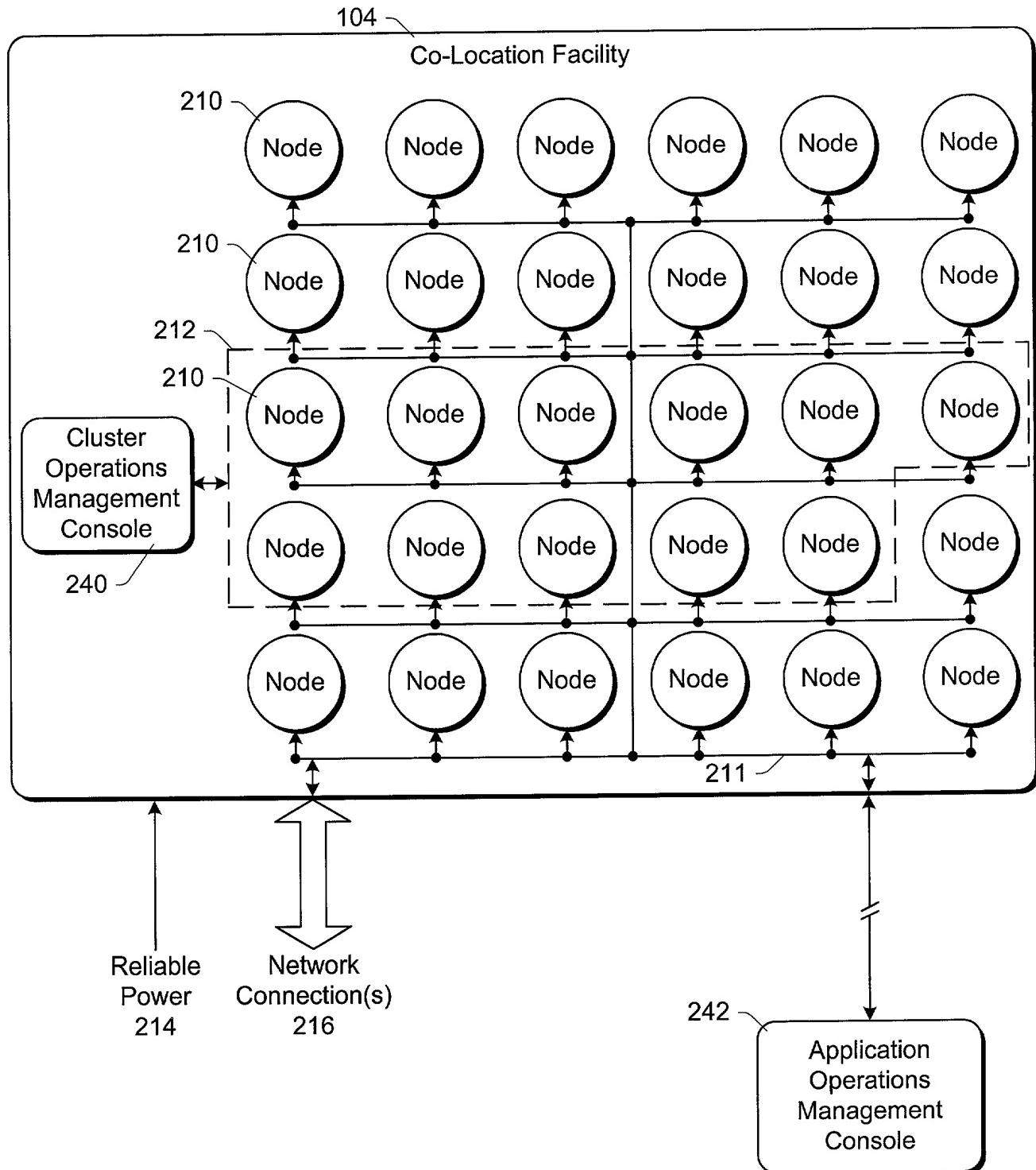


Fig. 3

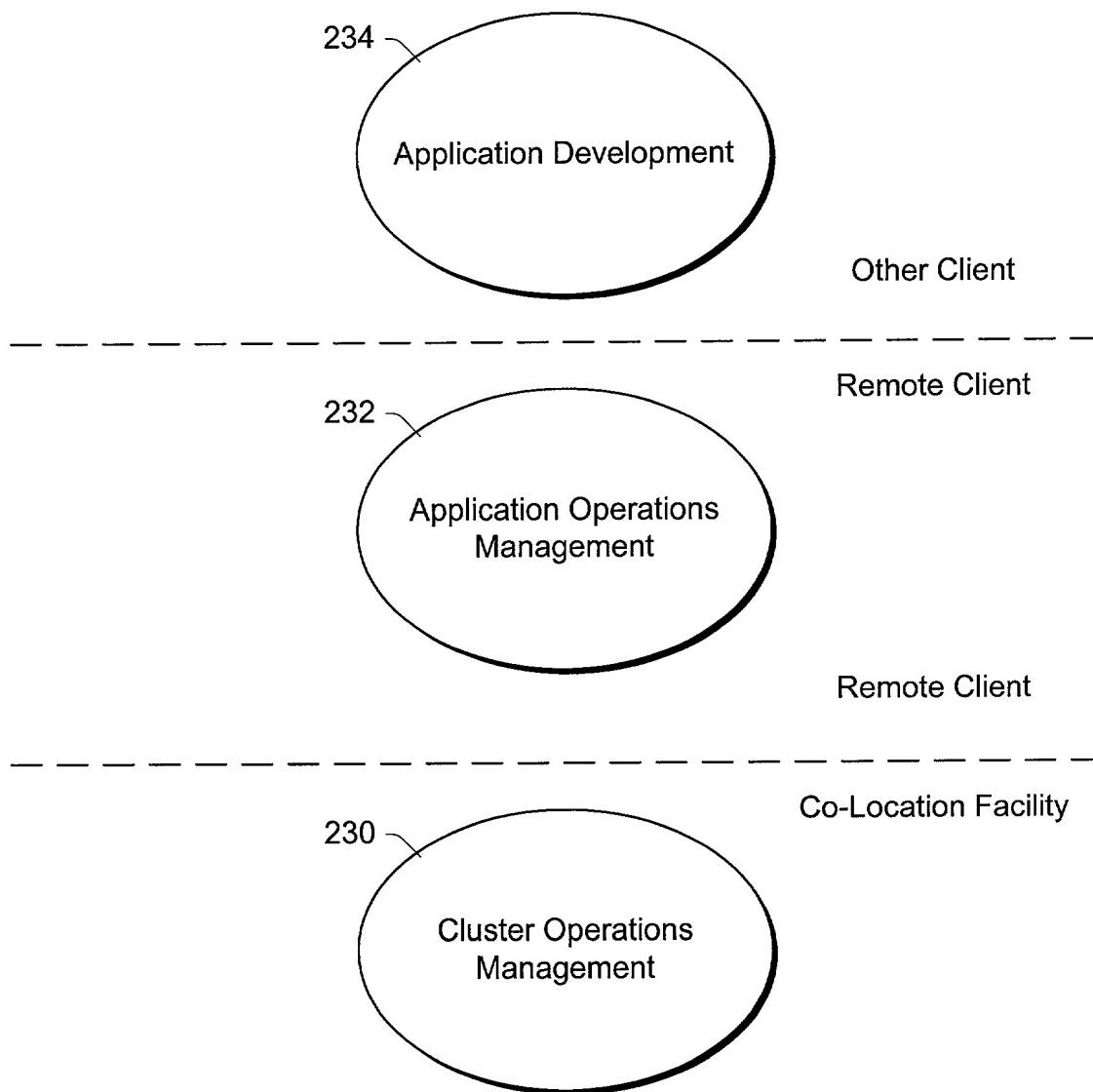


Fig. 4

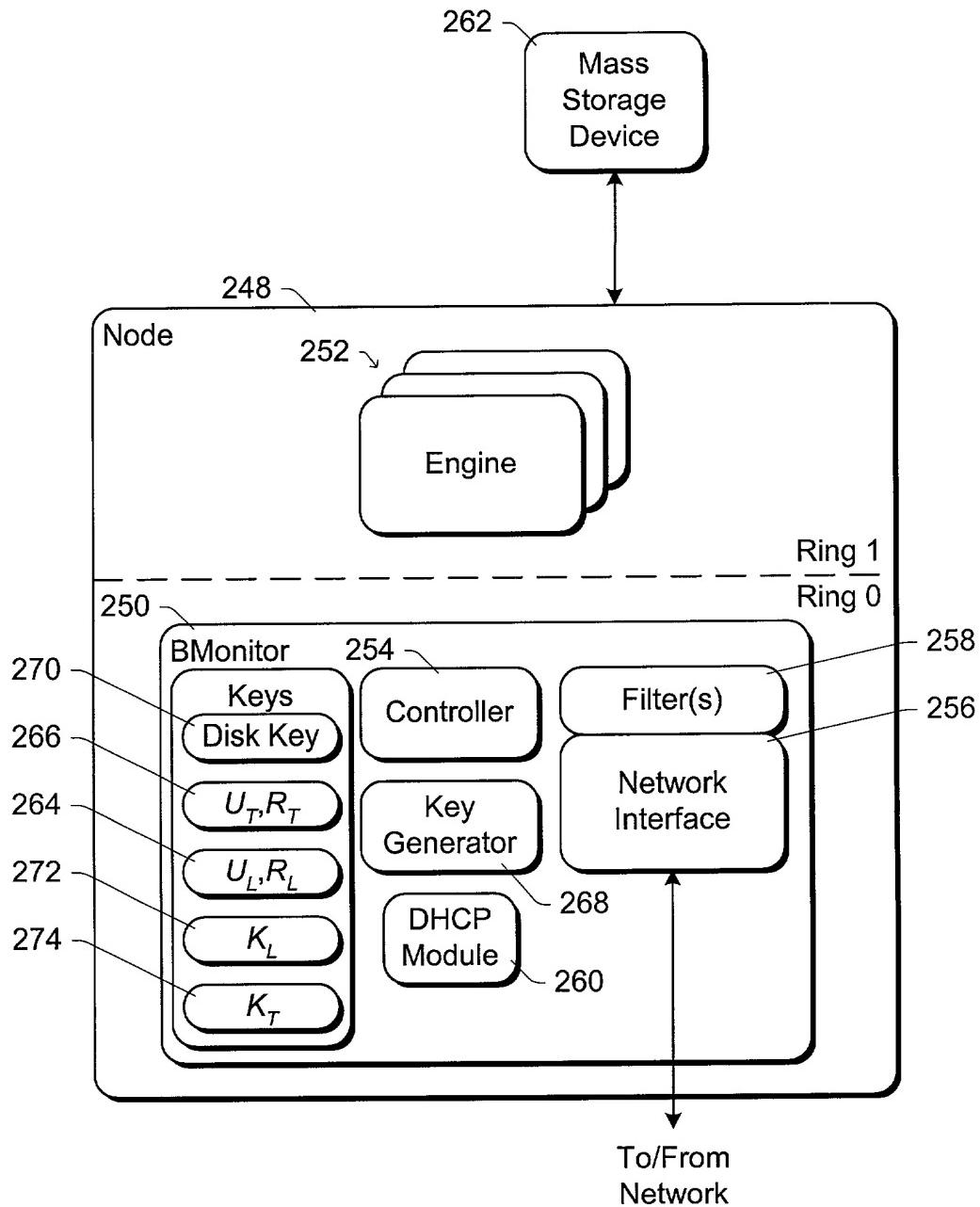


Fig. 5

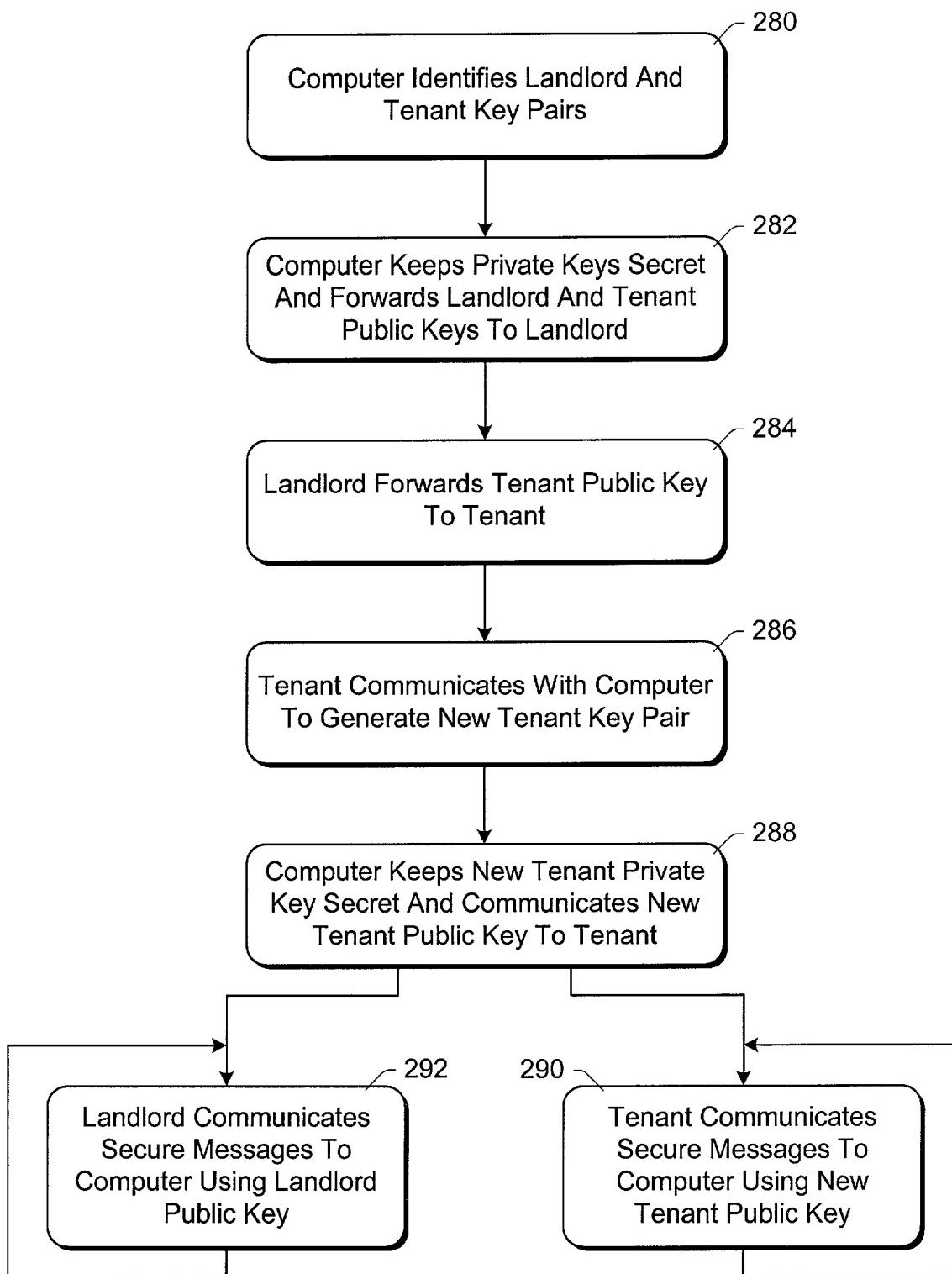


Fig. 6

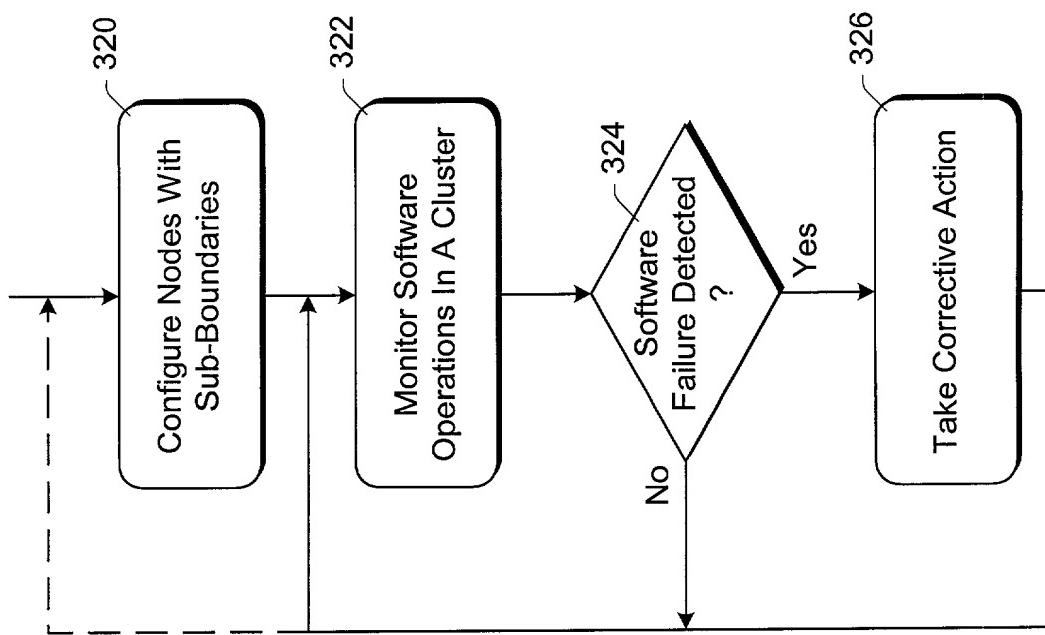


Fig. 8

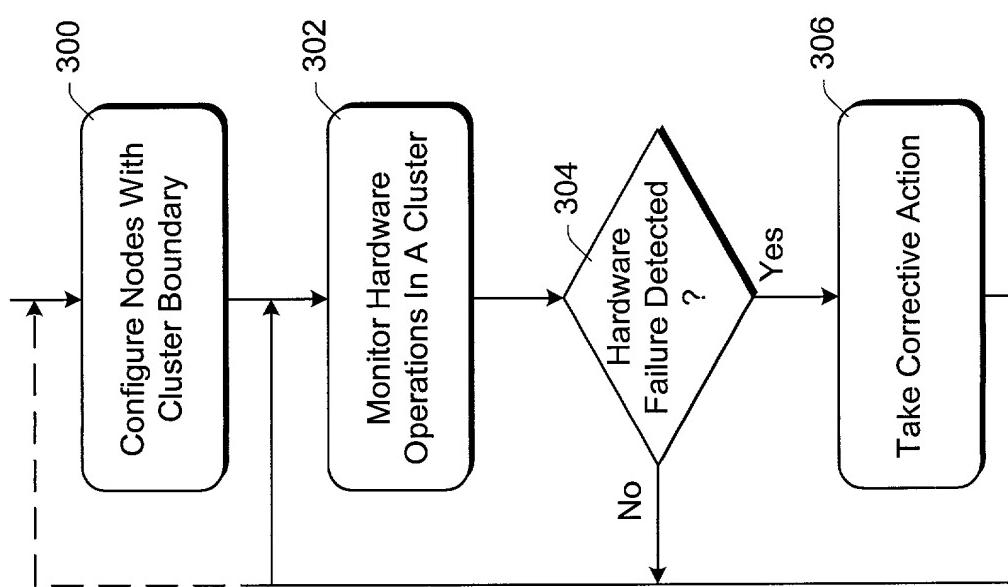


Fig. 7